J. Appl. Environ. Biol. Sci., 4(3)68-71, 2014

© 2014, TextRoad Publication

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Detecting Spammers

Sahar Bahramzadeh[1], Mehdi Hosseinzadeh[2]

[1]Department of Software Engineering, Science and Research Branch, Azad Islamic University, Saveh, Iran
[2]Department of Software Engineering, Science and Research Branch, Azad Islamic University, Tehran, Iran

Received: August 23 2013
Accepted: October 28 2013

ABSTRACT

Spam e-mails dedicated the large amount of bandwidth to themselves. Besides wasting bandwidth, these Spam have led to security threats of computer systems in network. Because most of the senders - known as spammers - appertain to botnet infected networks; therefore, in this research some measures were considered to identify spammer in internal network both based on analysis of e-mail content and analysis of traffic characteristics of e-mail packets. Thus, we can keep our network safe from the possible damages and also can have an efficient use of bandwidth of network. The result of assessment of this research indicates that our approach has higher accuracy and speed in detecting spammer than the previous systems.

KEYWORDS: Detecting Spammers, Approaches for Spam Bots Detection.

I. INTRODUCTION

Today, email is considered as a main part of human living, but by development of internet network and using of email for increased speed of data sharing to users in different subjects and despites the usefulness of this technology, we can see mass emails from some individuals and organization on daily based to some extent that they comprise high percentage of total email traffics. On 2010, the rate of spam emails included 88.5% of total emails sent in any day, i.e. up to 61.6 billion emails in each day, included spam emails [1].

Spam emails comprise the biggest problem of emails; because they allocate high size of bandwidth to themselves. In recent years, the security risk and threat of spams has been increased as well. Spams risks are because of high percentage of spam emails sent by botnets. On 2011 for example, in some months, botnets were responsible for sending about 95% of spam emails [2].

There have been considered different approach in spammer detecting systems designed since, for detecting the spam bots. Some of such systems detect spammer system relying on content-based methods in which there has been used an anti-spam filter like spamassassin. Some other ones are being detected relying on traffic specifications belonged to spammers that there is no analysis applied on the text of email. In the third approach, called "Hybrid Approach", one can use both previous approaches as two efficient and complementary factors to detect the spammers [3].

This paper introduces a new system. This is a developed version of SPOT system [4] performing based on network edge and detects spam bots on online based, by analysis of output email. Network edge means connecting point of a network or organization with other networks. This system has used on hybrid approach; meaning besides considering the content of email, we mainly analyze traffic specifications of email packets. We believe that this system is an evolved version of SPOT because it lacks the limitations observed in SPOT. Benefiting from combined approaches, i.e. using anti-spam filter as a content-based method and analyzing the traffic specifications of email as a non-content-based method, we tried to increase the rate of detecting the spam bot.

The remainder of the paper is organized as follows. Part 2 of this paper is allocated to related work in spammer detection area. Part 3 of this paper introduces designed system and pre-requirements needed for developing it. Part 4 indicates the results from this system evaluation and its performance result in a real network; and part 5 allocated to the limitations of study and finally, the conclusion is presented in part 6.

II. LITERATURES REVIEW

This part reviews the similar studies previously conducted and focused on some studies dealt with detecting compromised machines based on spammers' activity:

*** Corresponding Author:** Sahar Bahramzadeh, Department of Software Engineering, Science and Research Branch of Saveh, Iran
sahar.bahramzade@gmail.com

On 2006, Xie et al, William University provided DBspam for detecting the spam proxies. This functioned on-line and based on network edge. Spam washing was conducted by open proxies was an extended trick for hiding the actual source sending email spam [5]. Comparing with designed tools, this study aims to not only investigate spam proxies but also all spam bots in the network.

On 2007, Gu et al. developed applied program, BotHunter based on malware infection scenario. This tool is to some extent an IDS. It compares sequential steps of communication and executive operation of a machine with models previously detected based on full successful infection process of a malware and detects potential infected machines in the network [6]. Comparing to BotHunter stressing on detecting whole life cycle of a bot, system designed in this paper only focuses on detecting the spam activities of bots in the network; therefore, it has less processing complexities than BotHunter.

Two recent studies [7] and [8] conducted on 2008 aimed to detect general specifications of botnets such as their size and members belonged to them by analyzing emails received from a big mail server such as Hotmail. As indicated in a study [7], the methodology included introduction of AutoRE framework in order to extract and analysis polluted URLs embedded in the context of email and consequently clustering the emails according to such urls. And methodology indicated in study [8] included tracing the similar spam emails received from campaigns belonging to existing botnets. Generally, such analysis has been conducted on recorded traffic of mail server. Therefore, such designed tools may not be effective in online determination of spam bot in the network. On the other side, such approaches focus on better perception of general specifications of spam botnets than determining the spammer in an internal network. While this study aims to design a system for determining the spammers in an internal network on online based.

SPOT system has been designed on 2012 to detect spam bot in the internal network [4]. This system functions based on online email traffic analysis and based on network edge. The main drawback of this system is that being a spammer is only detected by relying on contents of emails sent by sender, while today's spammers arrange the content of their emails such not being detected by anti-spam filters. Therefore, under such conditions it is better to study the traffic specifications of email packets besides considering the contents of email using spamassassin to such extent that in some cases, we can detect the presence of spammer without benefiting from email content analysis and only by using such properties. In SPOT system also the messages sent are analyzed randomly and independent from each other, but in this system, we concentrated on abundant and similar messages sent by a machine as well. Comparing to SPOT that considered any sender as an end user, we normally consider email sender as an end user or an internal mail server.

III. A REVIEW ON OUR METHOD

This designed system analyzes output emails of network on port 25. In some networks, the traffic of port 25 is blocked by edge router [9], but it is unlike in most networks. Anyway, because our system is located before edge router, it can analyze the traffic of port 25 by receiving a copy of network traffic using TAP device and detect the spammers in that network.

Normally, email senders in a network are either final user or internal mail servers. As we know, there are also email senders or end user (c) and or internal mail servers (S) in our network. After our internal mail servers being detected, their ip is considered as white list. Even though such mail servers send many spam emails, because their actual sender machine couldn't be traced, therefore, such emails will not be analyzed and they will be considered as secure. Our system doesn't know ip address of mail servers before, but by running the application program it will periodically detect them and add to its database.

As mentioned above, detected spammer systems are among end users systems in the network. According to our assumptions, an end user machine in this network is either normal or compromised and this paper focused on detecting the compromised machines called spammer.

As in introduction, anti-spam filter, spamassassin, using in our approach is acted as a content-based method and we considered the analysis of traffic properties of senders as a non-content method (based on behavior).

A system is spam bot in our method, according to content-based approach if:

A. Detecting Spammers In Network

1) *Sending 3 emails with spam content:* According to achievements of SPOT study, if a system sends three emails that according to spamassassin filter the content of those emails are detected as spam, then the sender is spammer [4].

In addition, there have been used other behavioral properties in this system for detecting the spammers. For this reason, a system according to behavior-based approache is a spam bot in this system if comes with one of the following properties:

2) *Sending email not in allowed time:* In this system, we defined a text file called ip time in which we determined the permissible time for sending email based on any ip.

Therefore, if ip belonging to a subnet sends email not in the allowed time is called an spammer. The reason for this idea is that some spam bots are systems belonging to a specific organization or company with a normal behavior during the day, but in some hours during the night with no user under such machines, they force by botnet control centers to send emails. Hence, if we permit the ips belonging to such subnets to send email only in their normal period, by such way we can detect some spam botnets in different subnets. Some users, of course like home users are allowed to send email 24 hours a day.

3) Sending N emails from a machine under conditions with 70% different senders: In SPOT systems, the frequency of emails sent by a machine couldn't be considered. We know that spammers mostly intend to send similar and many emails. Results of our studies indicate that under normal state, emails sent by a machine are sent in an interval not very long maximum with 3 or 4 different IDs. But if for example 7 out of 10 emails sent by different IDs, it is considered as an anomaly and sender is called as a spammer.

4) Sending email under conditions that content of sender field is different in the header and body of email: Under normal conditions, the content of sender field in the header must be similar to its content in the body of email [10]. Otherwise, such anomaly in the behavior of sender as shown in figure 1 indicates it is a spammer.

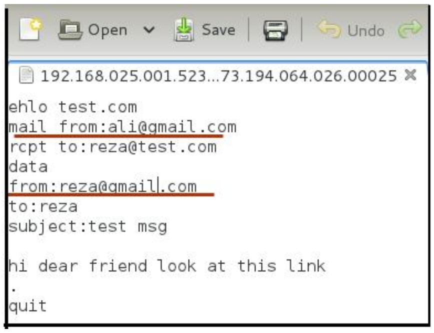
B. Detecting Email Sending Internal Mail Servers in the Network

Because email senders in the network might be our permissible internal mail servers, so it is necessary to make some arrangements to detect them and because in this study, we put the internal mail servers in white list, therefore, we will not analyze emails sent by them and consider them as secure. Generally, in a domain, some machines are only permitted to send email from that domain. Ip of such machines has been recorded in SPF record belonged to that domain in DNS server. On the other side, in any domain, only some machines are allowed to receive email from that domain. Ip of these machines in the MX record belonging to that domain has been recorded in DNS server. Therefore, in such domains, MX recording machines are responsible for receiving email and SPF recording machines are responsible for sending email. Therefore, if email sender IP is among SPF machines, mail sender is a legal server. In some other domains with no SPF record, MX record machines are responsible both for sending and receiving the email. Therefore, in such domains, if Sender's System IP belongs to MX record set, it is a legal mail server [11].

IV. LIMITATIONS AND FUTURE WORK

To facilitate conducting this study, we didn't consider emails sent by dynamic IPs and only analyzed emails belonging to static IPs. But as you know, some part of spam emails are sent by devices with dynamic address. Therefore, detecting spam zombies using dynamic IPs will be studied in a future work.

Because email packets are sent under different protocols and naturally on different ports as well, we aimed to investigate only the traffic of email under SMTP protocol on port 25. Then, there are certainly other spammers sent their emails under other protocols on ports 465 and or 578 and development of OUR SYSTEM for activity on such ports will be studied in a future work.



```
ehlo test.com
mail from:ali@gmail.com
rcpt to:reza@test.com
data
from:reza@gmail.com
to:reza
subject:test msg

hi dear friend look at this link
.
quit
```

Fig. 1. Different sender fields in header and body

V. CONCLUSION

In this study, we developed a new system to detect spammers in internal network. It is a developed version of SPOT and by analyzing the output emails of network edge, online, it detects spam zombies. In order for reducing the determination error in this alternative, this system analysis has been conducted based on both content of email and their traffic properties.

Studies indicate that this system has a proper and more accurate performance for detecting the spammers in the network and can be used as a complementary tool in Unified Threats Management Systems(UTM).

REFERENCES

[1] Analysis of Spam Activity Trends, 2011, Available: http://www.symantec.com/threatreport/topic.jsp>?id=spam_fraud_activity_trends&aid=analysis_of_spam_activity_trends [Accessed: March. 01, 2012].

[2] Percentage of spam sent from botnets, 2011. Available: http://www.symantec.com/threatreport/topic.jsp?id=spam_fraud_activity_trends&aid=analysis_of_spam_delivered_by_botnets [Accessed: March. 02, 2012].

[3] A. Ramachandran, and N. Feamster, "Understanding the Network-Level Behavior of Spammers", in *Proc. ACM SIGCOMM, NY*, USA. 2006, pp 291-302.

[4] Zhenhai Duan, Peng Chen, Fernando Sanchez, Yingfei Dong, Mary Stephenson, James Barker, "Detecting Spam Zombies by Monitoring Outgoing Messages" ,2012, *IEEE Trans. Inform. Theory*, vol 9,pp 198-210.

[5] M. Xie, H. Yin, and H. Wang, "An effective defense against email spam laundering" in *ACM Conference on Computer and Communications Security,* , Alexandria, VA, October 30 -November3 2006.

[6] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee, "Bothunter: Detecting malware infection through ids-driven dialog correlation," *in Proc. 16th USENIX Security Symposium*, Boston, MA, Aug. 2007.

[7] Y. Xie, F. Xu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, "Spamming botnets: Signatures and characteristics," in *Proc. ACM SIGCOMM*, Seattle, WA, Aug. 2008.

[8] L. Zhuang, J. Dunagan, D. R. Simon, H. J. Wang, I. Osipkov, G. Hulten, and J. D. Tygar, "Characterizing botnets from email spam records" in *Proc. of 1st Usenix Workshop on Large Scale Exploits and Emergent Threats*, San Francisco, CA, Apr. 2008.

[9] S. Linford, "Increasing spam threat from proxy hijacking," http://www.spamhaus.org/news.lasso?article=156.

[10] J. Klensin, "Simple Mail Transfer Protocol," RFC 5321, Oct. 2008.

[11] Fernando Sanchez and Zhenhai Duan and Yingfei Dong, "Blocking Spam By Separating End-User Machines from Legitimate Mail Server Machines" , *Proceedings of 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS),* Redmond, Perth, Australia, 2011.